

High Performance Memory Architecture for Vector Processors (ベクトルプロセッサのための 高性能メモリアーキテクチャに関する研究)

著者	撫佐 昭裕
号	14
学位授与番号	433
URL	http://hdl.handle.net/10097/42618

氏名（本籍地）	撫佐 昭裕
学位の種類	博士（情報科学）
学位記番号	情博第433号
学位授与年月日	平成21年3月25日
学位授与の要件	学位規則第4条第1項該当
研究科、専攻	東北大学大学院情報科学研究科（博士課程）情報基礎科学専攻
学位論文題目	High Performance Memory Architecture for Vector Processors (ベクトルプロセッサのための高性能メモリアーキテクチャに関する研究)
論文審査委員	(主査) 東北大学教授 小林 広明 東北大学教授 山本 悟 東北大学教授 中橋 和博 (工学研究科) 東北大学准教授 滝沢 寛之

論文内容の要旨

第1章 緒論

科学技術研究においてコンピュータシミュレーションの果たす役割は大きく、理論研究や実験では不可能であった研究が可能となった。科学技術計算の分野では主にベクトル型スーパーコンピュータと汎用マイクロプロセッサによるスカラ型コンピュータがあり、近年、米国を中心としてスカラ型コンピュータが主流となってきている。しかし、スカラ型コンピュータにおいては、実シミュレーションプログラムにおける、コンピュータの理論性能と実効性能の著しい乖離が存在している。一方、ベクトル型スーパーコンピュータは、実アプリケーションプログラムにおいて実行効率が40%を超える、高い実効性能を有していることが示されている。

これらコンピュータの実効性能はメモリシステムに依存している。スカラ型コンピュータはキャッシュメモリによる階層型メモリアーキテクチャを、また、ベクトル型スーパーコンピュータはインターリーブ方式によるメモリアーキテクチャを採用し、CPUに対し高速にメモリデータを転送できるようになっている。しかし、メモリデータ転送性能と演算性能のギャップは年々拡大し、メモリシステムのデータ転送性能がボトルネックとなり、CPU性能を十分に発揮できないというMemory Wall問題が存在している。特に、ベクトル型スーパーコンピュータにおいてMemory Wall問題により、今まで実証されてきた高い実効性能が維持できなくなる問題に直面している。

本論文では、ベクトル型スーパーコンピュータの性能をメモリ性能の観点で議論し、メモリバンド幅が性能にもっとも重要な要素であることを示す。また、ベクトル型スーパーコンピュータのメモリバンド幅が将来低下することを想定し、高性能なメモリアーキテクチャ（ベクトルキャッシュ機構）を提案し、その有効性を示すとともにベクトルキャッシュ機構の特性を明らかにする。

第2章 高実効性能スーパーコンピュータのためのメモリ性能

本章では、実アプリケーションプログラムを用いて、ベクトル型スーパーコンピュータとスカラ型コンピュータの実効性能についてメモリ性能の観点から実機を用いた評価を行い、メモリアーキテクチャと実行効率（ピーク性能に対する実効性能の比）の関係を明らかにする。本論文では、メモリバンド幅とCPU

性能の比 B/FLOP を用いて議論を行う ($B/FLOP = \text{メモリバンド幅 (B/s)} / \text{演算性能 (flop/s)}$).

性能評価には、ベクトル型スーパーコンピュータとして NEC SX-7 と SX-8 (各 4B/FLOP) を、スカラ型コンピュータとして NEC TX-7 と SGI Altix3700 (各 1B/FLOP) を用いる。評価用プログラムとして、電磁解析、熱流体及び地球科学分野から 5 本の実シミュレーションプログラムを用いる。

各システムにおける 1 CPU の実行効率は、ベクトル型スーパーコンピュータで 40%以上を達成しているが、スカラ型コンピュータでは最大でも 10%を超える程度である。これは、各システムの実効性能がメモリシステムに大きく依存しているからである。ベクトル型スーパーコンピュータが採用しているインターリーブ方式は、多数のバンクによってメモリレイテンシを隠蔽するとともに、ベクトルパイプラインによる演算とメモリアクセスをオーバーラップさせることができる。このことにより、5 本のプログラムにおいて、演算に隠蔽できなかったメモリアクセス時間の実行時間に対する割合は 30%以下であった。一方、階層型メモリシステムのスカラ型コンピュータでは、演算に隠蔽できなかったメモリアクセス時間の割合は 80%以上になっている。このことから、ベクトル型スーパーコンピュータが採用しているインターリーブ方式は、実アプリケーションプログラムを高速に処理するのに適したメモリシステムであるが、本評価から示された。

さらに、ベクトル型スーパーコンピュータのメモリシステムにおいて、メモリバンド幅比 (B/FLOP) と実行効率の関係を調査した。ベクトル型スーパーコンピュータ SX-7 のメモリバンド幅比を 1/2, 1/4 に絞り、実行効率がどのように変化するかを評価した。その結果、メモリバンド幅比を 1/2, 1/4 にすることにより、演算に隠蔽できないメモリアクセス時間が 2 倍、4 倍と長くなり、メモリバンド幅比に比例して実行効率が低下した。このことより、ベクトル型スーパーコンピュータはメモリバンド幅比によって性能が維持されていることが明らかになった。

以上より、実アプリケーションプログラムにおいてベクトル型スーパーコンピュータの性能はスカラ型コンピュータより高く、この性能はメモリシステムによることを示した。また、ベクトル型スーパーコンピュータの実行性能はメモリバンド幅比に強く依存し、高い実行効率を維持するためには、4B/FLOP のメモリバンド幅比が必要であることを明らかにした。

第 3 章 ベクトルアーキテクチャのためのオンチップキャッシュ

本章では、将来、Memory Wall 問題によってベクトル型スーパーコンピュータのメモリバンド幅比が低下した場合を想定し、ベクトル型スーパーコンピュータの性能を維持する高性能メモリアーキテクチャを提案し、その有効性を実証する。

高性能メモリアーキテクチャの目的は、メインメモリとベクトルプロセッサ間のデータ転送を削減し、メモリバンド幅を有効に使用することである。一般に科学技術計算は、差分方程式を代表例としてプログラム内にデータ局所性を有している。その局所性を最大限利用するため、高性能メモリアーキテクチャとしてオンチップキャッシュ機構 (ベクトルキャッシュ機構) を提案する。ベクトルキャッシュ機構は、ベクトルプロセッサに転送されたデータを再利用することによって、メインメモリとベクトルプロセッサ間のデータ転送を削減させるものである。また、ベクトルキャッシュにはメインメモリとベクトルレジスタ間にバイパス機構を設け、ベクトルレジスタにおけるメモリバンド幅比を 4B/FLOP に維持するものである。さらに、ベクトルキャッシュには、Miss Status Handling Register (MSHR) 機構とプリフェッチ機構を搭載する。MSHR 機構は、ベクトルデータのロード中ステータスを管理し、重複した連続メモリアクセスを防止し、メインメモリとベクトルプロセッサ間のデータ転送の負荷を軽減するものである。また、プリフェッチ機構は、演算に必要なデータを事前にロードする機能であり、メモリ負荷を分散するとともにメモリレイテンシを隠蔽するものである。また、プリフェッチ機構は MSHR 機構と同様に重複データの転送を防止することができる。

ベクトルキャッシュ機構の評価として、ベクトルプロセッサの動作をシミュレートするためのレジスタトランスファレベルのタイミングシミュレータを開発し、2 章で使用した 5 本の実アプリケーションプログラムを用いた。

本評価によって、ベクトルキャッシュ機能は、メモリバンド幅比 2B/FLOP と 1B/FLOP において低下し

たベクトルレジスタにおけるメモリバンド幅を向上させることを明らかにした。そして、2B/FLOP におけるベクトルプロセッサの実行効率が、ベクトルキャッシュ機構によって 4B/FLOP システムの実行効率に対して 21%から 96%まで改善することを明らかにした。特にキャッシュヒット率が 50%以上の場合、メモリバンド幅比 2B/FLOP のシステムでは、4B/FLOP システムと同等な実行効率が得られている。これは、バイパス機能によって、ベクトルキャッシュとメインメモリの両方から同時にデータを供給していることによるものである。また、MSHR 機構は、2B/FLOP システムにおいて、さらに 5%から 25%の性能改善を行い、プリフェッチ機構は 20%から 30%の性能改善を行うことも明らかになった。

以上より、本提案の高性能メモリアーキテクチャ（ベクトルキャッシュ機構）は、将来、メモリバンド幅比が低下したベクトル型スーパーコンピュータにおいて、実効性能を維持する有効な機能であることを明らかにした。

第4章 チップマルチベクトルプロセッサのための共有キャッシュ機構

ベクトルプロセッサは、近い将来、マルチコアアーキテクチャが採用され、今以上にメモリバンド幅比の低下が予測されている。本章では、チップマルチベクトルプロセッサ(CMVP)におけるメモリバンド幅比の低下を補完する機構として前章で示したキャッシュ機構を提案し、その有効性を明らかにする。

マルチスレッド化された科学技術計算の多くはスレッド間にデータ局所性を有している。この局所性を有効に活用するため、コア間でベクトルキャッシュを共有する共有型キャッシュ機構を提案する。そして、前章で有効性を示したバイパス機構と MSHR 機能を採用する。また、CMVP として 4 コアのベクトルコアを有する構成を想定し、2 章で使用した 5 本の実アプリケーションプログラムで評価を行う。

本評価によって、ベクトルキャッシュを有しない CMVP においては、高い並列性能を実現するためにはメモリバンド幅比 4B/FLOP が必要であることを明らかにした。また、ベクトルキャッシュを有する CMVP では、メモリバンド幅比 2B/FLOP において、4B/FLOP のシステムと同等な並列性能が得られることが判明した。これは、バイパス機構によりメモリとキャッシュ機構の両方からデータ供給されることによるものである。また、MSHR 機構は、コア間の同一データのアクセスにおいて有効な機構であり、4 コアにおける実効性能を 17% から 50% 向上させることを明らかにした。

以上より、本提案の共有型ベクトルキャッシュ機構は、将来採用される CMVP において、並列性能を維持する有効な機構であることを明らかにした。

第5章 結論

本論文では、ベクトル型スーパーコンピュータの実効性能の高さを実アプリケーションプログラムによって示し、その性能はメモリバンド幅比(B/FLOP)がもっとも重要な要素であり、4B/FLOP が必要であることを明らかにした。また、ベクトル型スーパーコンピュータのメモリバンド幅比が将来低下することを想定し、高性能なメモリアーキテクチャ（ベクトルキャッシュ機構）の提案を行った。ベクトルキャッシュ機構には、メモリとベクトルレジスタ間にバイパス機構を設け、さらに MSHR 機能とプリフェッチ機構を採用した。本論文では、ベクトルキャッシュ機能の有効性を示すとともにベクトルキャッシュ機構の特性を明らかにし、将来のベクトル型スーパーコンピュータにおける有効なメモリアーキテクチャであることを明らかにした。

論文審査結果の要旨

先端科学分野においてシミュレーションは、実験、理論と並ぶ第3の手法としてますます重要になっている。ベクトルプロセッサは、先端科学分野で求められる多様なシミュレーションの大規模科学計算を高い実効性能で処理できることから、スーパーコンピュータの主要な構成要素になっているが、多数の演算器のオンチップ化やマルチコア化によるプロセッサ性能の著しい向上に対しメモリ性能が伴わず、ベクトルプロセッサ向け高性能メモリアーキテクチャの確立が急務となっている。本論文は、実際のシミュレーションコードにおけるベクトルプロセッサのメモリ操作を詳細に分析し、効率よいデータ供給を実現する高性能メモリアーキテクチャをハードウェアとソフトウェアの両面から論じたもので、全編5章よりなる。

第1章は緒論である。

第2章では、ベクトルプロセッサの実効性能とメモリ性能の関係を明らかにするために、計算科学分野の実用的なシミュレーションコードを用いて、ベクトル型スーパーコンピュータとスカラ型スーパーコンピュータの性能評価を行っている。そして、メモリ性能と演算性能の比である B/F (Bytes/Flop) 値が、実効性能を評価する上で重要な指標になることを明らかにし、ベクトルプロセッサの高い実効性能を維持するには、4B/F 以上のデータ供給能力が必要であることを導いている。これは高性能メモリアーキテクチャを設計するための重要な指針である。

第3章では、レジスタレベルで 4B/F のデータ供給能力と冗長なメモリアクセスをハードウェアで検出する MSHR 機構を有するシングルベクトルコア向けキャッシュアーキテクチャを提案している。本キャッシュアーキテクチャにおけるデータ管理をソフトウェア制御可能とすることにより、局所性の高いデータのキャッシュ格納、および局所性の低いデータのキャッシュバイパスが可能となり、少ない容量で高いベクトルデータ供給能力を実現している。実シミュレーションコードを用いた詳細な評価実験から、オフチップメモリバンド幅が 2B/F であるシステム構成においても 4B/F に匹敵する実効性能が得られることを明らかにしている。これは、重要な成果である。

第4章では、第3章で提案したキャッシュアーキテクチャをマルチコアプロセッサ用に拡張し、性能評価を行っている。まず、マルチスレッド実行におけるベクトルデータ間の参照の局所性をシミュレーションコードの分析から明らかにし、次に、マルチコア間でベクトルキャッシュによりデータ共有能力を高めたチップマルチベクトルプロセッサアーキテクチャを提案している。実シミュレーションコードを用いた性能評価の結果、提案アーキテクチャは、オフチップメモリバンド幅の制限にもかかわらず、コア数に比例した性能向上が達成できることを明らかにしている。これらは、高性能スーパーコンピュータの要素技術として、ベクトルプロセッサのマルチコア化を促進する重要な成果である。

第5章は、本論文を総括し、結論としている。

以上要するに本論文は、次世代スーパーコンピュータ設計における高性能メモリアーキテクチャの重要な基礎を与えたもので、情報基礎科学および計算機科学の発展に寄与するところが少なくない。

よって本論文は、博士(情報科学)の学位論文として合格と認める。